Fakulti Teknologi Maklumat dan Sains Kuantitatif

Jilid 4, Bil. 1, 2002

JURNAL TEKNOLOGI MAKLUMAT DAN SAINS KUANTITATIF

Kandungan	Muka Surat

Universiti Teknologi MARA

Jurnal Tek. Maklumat & Sains Kuantitatif Jilid 4, Bil. 1, 2002 (45-57)

Pemodelan tuntutan insurans bagi perbelanjaan perubatan (kajian kes)

Noriszura Hj. Ismail dan Yeoh Sing Yee
Program Sains Aktuari, Pusat Pengajian Sains Matematik, FST, UKM Bangi 43600, Selangor

Abstrak
Kajian ini bertujuan untuk menyesuaikan suatu taburan berparameter terhadap sampel data tuntutan perbelanjaan perubatan. Sampel data yang diperoleh adalah salah satu produk insurans kesihatan daripada syarikat insurans yang terdapat di Malaysia. Sampel data ini bertempoh 30 bulan dan bermula dari Januari 1999 sehingga Jun 2001. Kaedah penganggaran kebolehjadian maksimum telah diguna untuk menganggar parameter-parameter taburan yang disuai. Penganggaran ini dilakukan dengan teknik pengoptimuman dengan bantuan perisian komputer. Ujian penerimaan model dilakukan dengan menggunakan ujian statistik Pearson khi kuasa-dua dan ujian statistik Kolmogorov-Smirnov. Pemilihan model yang terbaik pula dilakukan dengan menggunakan ujian nisbah kebolehjadian dan ujian Schwartz Bayesian.

Kata kunci: Insurans kesihatan; Kaedah kebolehjadian maksimum; Taburan berparameter; Tuntutan perbelanjaan perubatan.

Abstract
This study aims to fit a parametric distribution on the sample data of medical expenses claims. The sample is obtained from one of the health insurance products from one of the insurance companies in Malaysia. The claims were paid from January 1999 until June 2001, a period of 30 months. Maximum likelihood procedure was used to estimate the parameters of the fitted distributions. Optimization technique from the computer software was used to carry out the procedure. Pearson chi-square statistical test and Kolmogorov-Smirnov statistical test were used to test the model's acceptability. Likelihood ratio test and Schwartz Bayesian test were used to determine the best model for the sample.

Keywords: health insurance; maximum likelihood procedure; claims on medical expenses; parametric distribution.

1. Pengenalan

Insurans kesihatan adalah suatu bentuk perlindungan yang disediakan oleh sesebuah syarikat insurans kepada insud (pembeli insurans) untuk menyelesaikan masalah bayaran kos rawatan dan perubatan serta mengatasi masalah kehilangan pendapatan yang disebabkan oleh kemalangan atau penyakit. Biasanya premium bagi polisi insurans kesihatan adalah lebih tinggi berbanding dengan premium yang melindungi risiko lain.

e-mail: ni@pkrisc.cc.ukm.my ; singyee.yeoh@inginsurance,com.my

Perlindungan insurans kesihatan boleh dikategorikan kepada 2 kumpulan utama iaitu perlindungan perbelanjaan kesihatan dan perlindungan pendapatan hilang upaya (Treischmann et al. 2001). Perlindungan perbelanjaan kesihatan boleh dikelaskan kepada 3 jenis iaitu insurans kesihatan asas, insurans perubatan utama dan insurans perlindungan pelbagai.

Insurans kesihatan asas boleh dibahagikan kepada 3 jenis iaitu insurans hospital, insurans pembedahan dan insurans perbelanjaan perubatan biasa. Insurans hospital memberikan manfaat kepada perbelanjaan hospital, termasuklah bilik di hospital, yuran makmal, rawatan di hospital, bilik pembedahan dan bekalan ubat. Polisi insurans pembedahan pula memberi manfaat kepada yuran doktor pakar bedah, termasuklah untuk prosedur pembedahan sebagai pesakit luar atau dalam. Akhir sekali, insurans perbelanjaan perubatan biasa memberi manfaat kepada yuran servis doktor selain daripada pakar yang melakukan prosedur pembedahan.

Insurans perubatan utama pula melindungi semua jenis perbelanjaan yang dilindungi oleh ketiga-tiga jenis insurans hospital, insurans pembedahan dan insurans perbelanjaan perubatan biasa. Selalunya insurans ini mempunyai had bendungan yang lebih tinggi berbanding dengan ketiga-tiga jenis insurans ini.

Di Malaysia, kos perubatan dan caj rawatan yang dikenakan kepada pesakit semakin meningkat setiap tahun. Berdasarkan indeks harga pengguna, peratus perubahan indeks daripada tahun 1991 ke 1995 bagi perubatan dan kesihatan adalah sebanyak 16%. Peratus perubahan ini adalah yang ketiga tertinggi selepas makanan, 20.3%, dan pengangkutan, 17.4% (Jabatan Perangkaan Malaysia 1996). Peningkatan kos perubatan dan caj rawatan yang tinggi menimbulkan masalah kepada syarikat insurans dalam menentukan premium. Di samping peningkatan ini, terdapat juga sesetengah doktor dan hospital yang mengenakan bayaran yang lebih tinggi daripada yang sepatutnya (over-charging) dan/atau memanjangkan tempoh tinggal pesakit di hospital supaya lebih lama daripada yang sepatutnya (over-servicing). Pada tahun 1987 misalannya, terdapat syarikat insurans yang mengalami tuntutan yang merangkumi 66% daripada jumlah premium kasar tahunannya (Chee 1990).

Kajian terhadap taburan tuntutan perubatan adalah penting kerana model yang diperoleh boleh diguna untuk menerangkan proses kerugian yang dilindungi. Untuk memastikan agar premium yang dikenakan adalah tepat menurut sebarang prinsip pengiraan premium, pengiraan yang dilakukan mestilah berasaskan proses kerugian yang dilindungi. Secara praktis, proses kerugian yang sebenar adalah sukar untuk diketahui. Namun, penganggaran proses kerugian dengan kejituan yang tinggi boleh memberikan asas untuk mendapatkan premium yang tepat.

Terdapat banyak kajian yang telah dijalankan oleh penyelidik terdahulu tentang penyesuaian taburan berparameter terhadap sampel data tuntutan insurans. Hewitt dan Lefkowitz (1979) telah cuba menyesuaikan 5 jenis taburan, gamma, loggamma, lognormal, gamma dan loggamma dan gamma dan lognormal terhadap data kerugian dan mengira kesan inflasi ke atas deduktibel dan had bendungan kerugian. Hasilnya, model taburan kompaun gamma & loggamma memberikan penyesuaian yang lebih baik terhadap data berbanding dengan model taburan lognormal. Selain itu, kelima-lima model yang disuai boleh diguna untuk mengira kesan inflasi ke atas deduktibel dan had bendungan kerugian. Philbrick (1985) telah cuba menyesuaikan taburan Pareto terhadap data insurans kemalangan yang terpangkas dan tertapis dengan menggunakan kaedah momen dan kaedah pemadanan. Hasilnya, penyesuaian taburan Pareto adalah memuaskan dan histogram dan ogif juga dapat ditunjukkan. Brockett et al. (1995) telah cuba menyelesaikan masalah data berkumpulan yang menggunakan data sekunder dengan min dan median yang diketahui. Hasilnya, beliau telah menunjukkan prosedur untuk menjana penganggaran kebarangkalian ketumpatan dalam bentuk histogram.

2. Penganggaran Berparameter

Kajian ini bertujuan untuk menyesuaikan suatu taburan berparameter terhadap sampel data tuntutan perbelanjaan perubatan. Sampel data yang diperoleh adalah salah satu produk insurans kesihatan daripada syarikat insurans yang terdapat di Malaysia. Terdapat banyak kaedah penganggaran taburan berparameter yang boleh digunakan. Namun, kaedah kebolehjadian maksimum adalah kaedah penganggaran berparameter yang terbaik, sekurang-kurangnya berdasarkan sifat-sifat statistiknya. Kaedah ini mempunyai beberapa sifat yang tidak dikongsi oleh penganggar berparameter yang lain.

Di antara sifat-sifat ini adalah (Klugman et al. 1998):

(a) Mereka adalah penganggar yang tidak pincang secara asimptot.
(b) Berbanding dengan penganggar lain yang tertabur dengan normal secara asimptot, mereka mempunyai varians asimpot terkecil.
(c) Penganggar kebolehjadian maksimum bagi suatu fungsi berparameter adalah sama dengan penganggar kebolehjadian maksimum bagi parameter.
(d) Rumus untuk varians asimptot penganggar boleh diterbitkan.

Oleh itu, di dalam kajian ini kaedah kebolehjadian akan digunakan untuk mendapatkan suatu taburan berparameter yang dapat mewakili sampel data.

Bagi kaedah kebolehjadian maksimum, fungsi kebolehjadian, $L(\underline{\theta})$, untuk suatu set cerapan yang saling tidak bersandar adalah,

$$L(\underline{\theta}) = \prod_{j=1}^{n} L_j(\underline{\theta})$$

dengan,

$L_j(\underline{\theta})$ =sumbangan cerapan ke-j kepada kebolehjadian.

Jika cerapan ke-j adalah suatu kejadian dengan kebarangkalian positif (seperti taburan diskret atau selang), sumbangan adalah kebarangkalian. Jika cerapan ke-j adalah nilai dari taburan selanjar, sumbangan adalah fungsi ketumpatan pada nilai tersebut.

Bagi data individu, fungsi kebolehjadian adalah,

$$L(\underline{\theta}) = \prod_{j=1}^{n} f(x_j \mid \underline{\theta})$$

dengan,

$f(x_j \mid \underline{\theta})$ = fungsi ketumpatan bagi x_j bersyaratkan parameter $\underline{\theta}$

Prosedur kebolehjadian maksimum adalah untuk memaksimumkan $L(\underline{\theta})$. Prosedur ini adalah sama dengan mendapatkan terbitan pertama logaritma $L(\underline{\theta})$ terhadap parameter $\underline{\theta} = \{\theta_1, \theta_2, \ldots.., \theta_k\}$ dan menyamakan persamaan yang diperoleh dengan sifar.

Untuk kajian ini, 15 model daripada pelbagai taburan telah dipertimbangkan untuk dijadikan sebagai model bagi sampel data. Taburan-taburan ini adalah taburan 1-parameter iaitu eksponen dan inverse eksponen, taburan 2-parameter iaitu Pareto, inverse Pareto, loglogistik, paralogistik, inverse paralogistik, gamma, inverse gamma, Weibull dan inverse Weibull, taburan 3-parameter iaitu Pareto teritlak, Burr, dan inverse Burr dan taburan 4-parameter iaitu beta terubahsuai.

Sebagai contoh, model taburan eksponen akan digunakan. Fungsi ketumpatan taburan eksponen dengan $\theta_1 = \theta$ adalah,

$$f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, x > 0, \theta > 0$$

Oleh itu, fungsi kebolehjadian adalah,

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \frac{1}{\theta^n}e^{-\frac{\sum_{i=1}^{n} x_i}{n}}$$

dan logaritma kebolehjadian pula adalah,

$$\ln L(\theta) = -\frac{\sum_{i=1}^{n} x_i}{\theta} - n\ln\theta$$

Terbitan separa logaritma kebolehjadian terhadap θ adalah,

$$\frac{\partial}{\partial\theta}\ln L(\theta) = \frac{\sum_{i=1}^{n} x_i}{\theta^2} - \frac{n}{\theta} = 0$$

Dengan menyelesaikan persamaan ini, anggaran untuk parameter θ ialah,

$$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Namun, hanya beberapa taburan sahaja yang anggaran parameternya dapat diselesaikan dengan menggunakan persamaan terbitan pertama $\ln L(\underline{\theta})$ terhadap $\underline{\theta}$. Parameter $\underline{\theta}$ selalunya harus diselesaikan dengan kaedah lelaran Newton-Raphson. Oleh itu, bagi kajian ini anggaran parameter $\hat{\underline{\theta}}$ diselesaikan dengan teknik pengoptimuman dengan bantuan perisian komputer. Sebelum proses lelaran dilakukan, nilai awalan harus diberi. Kemudian, dengan memaksimumkan $\ln L(\underline{\theta})$, anggaran parameter $\hat{\underline{\theta}}$ dapat diselesaikan.

3. Ujian Penerimaan Model

Ujian kebagusan penyesuaian model selalunya diguna untuk menentukan sama ada model yang diperoleh diterima atau ditolak. Ujian ini menguji sama ada taburan bagi sampel diambil dari suatu taburan berparameter yang diketahui. Biarkan $F(x;\underline{\theta})$ sebagai fungsi ketumpatan taburan berparameter yang tersuai. Jika $F_X(x)$ adalah fungsi ketumpatan populasi yang sebenar, kita ingin menguji hipotesis,

$$H_0 : F_X(x) = F(x;\underline{\theta})$$
$$H_1 : F_X(x) \neq F(x;\underline{\theta})$$

Ujian kebagusan penyesuaian yang sering diguna untuk menguji kesesuaian model berparameter adalah ujian statistik Pearson khi kuasa-dua dan ujian statistik Kolmogorov-Smirnov.

3.1 Ujian statistik Pearson khi kuasa-dua

Untuk ujian ini, sampel rawak $X_1, X_2, \ldots, X_n$ yang tidak bersandar dicerap dan dibahagi kepada r kumpulan, $(c_0,c_1], (c_1,c_2], \ldots, (c_{r-1},c_r]$. Kekerapan cerapan untuk setiap kumpulan dikira dan nilai-nilainya adalah $n_1, n_2, \ldots, n_r$. Di bawah hipotesis nol, kebarangkalian bagi kumpulan ke-j, p_j, adalah,

$$p_j = F(c_j) - F(c_{j-1}), j = 1,2,\ldots,r$$

dan jangkaan bilangan cerapan dalam kumpulan ke-j, E_j, adalah,

$$E_j = np_j, j = 1,2,\ldots,r$$

dengan,

n = bilangan sampel data.

Rumus untuk ujian statistik Pearson adalah,

$$\chi^2 = \sum_{j=1}^{r} \frac{(n_j - E_j)^2}{E_j}$$

Jika n besar, χ^2 dianggar bertaburan khi kuasa-dua dengan darjah kebebasan, v=r-k-1 (r ialah bilangan kumpulan dan k ialah bilangan anggaran parameter taburan). Oleh itu, jika χ^2 adalah kecil, $F(x;\underline{\theta})$ diterima sebagai model yang munasabah. Untuk memastikan agar ujian ini dapat dipakai, setiap nilai E_j, j=1,2,.....,r, mestilah lebih daripada 5. Jika nilai E_j adalah kurang daripada 5, batas atas perlu dikembang atau lebih kumpulan perlu digabung untuk membentuk satu kumpulan dengan nilai E_j yang melebihi 5. Model taburan yang mempunyai nilai χ_{kira}^2 yang terkecil merupakan model taburan kerugian yang terbaik.

3.2 Ujian statistik Kolmogorov-Smirnov

Ujian statistik Pearson tidak sesuai diguna jika bilangan cerapan adalah kecil. Sebagai alternatif, ujian statistik Kolmogorov-Smirnov boleh diguna kerana kaedah ini menggunakan sampel data individu. Untuk kaedah ini, biarkan $F_n(x)$ sebagai fungsi longgokan empirik. Rumusnya ialah,

$$F_n(x) = \frac{bilangan X_i \le x}{n}$$

Ujian statistik Kolmogorov-Smirnov, D_n, ditakrifkan sebagai perbezaan mutlak maksimum di antara fungsi longgokan tersuai, $F(x;\underline{\theta})$, dengan fungsi longgokan empirik, $F_n(x)$. Rumusnya adalah,

$$D_n = maks_x \, | F_n(x) - F(x;\underline{\theta}) |$$

$F_n(x)$ adalah tidak selanjar dan mempunyai lompatan pada setiap titik sampel cerapan, x_1, x_2, ..., x_n. Jika $F(x;\underline{\theta})$ adalah selanjar, nilai maksimum sisihan mesti terjadi sebaik sebelum atau selepas lompatan pada $F_n(x)$. Oleh itu, rumus untuk D_n adalah,

$$D_n = maks_n\left\{ \left|F_n(x_i^-) - F(x;\underline{\theta})\right|, \left|F_n(x_i) - F(x;\underline{\theta})\right| \right\}$$

dengan,

$\{x_i\}$ = tertib sampel $x_1 \le x_2 \le ..... \le x_n$.

$F_n(x_i)$ = fungsi longgokan empirik pada x_i

$F_n(x_i^-)$ = fungsi longgokan empirik pada pada x_{i-1}

Di bawah H_0, D_n mempunyai taburan yang tidak berapa dikenali (Hogg & Klugman 1984). Namun, jadual untuk taburan D_n boleh diperoleh dari beberapa buku statistik.

4. Ujian Pemilihan Model

Ujian untuk pemilihan model adalah penting kerana jika beberapa model dengan parameter yang berbeza telah diterima, pemilihan model yang terbaik perlu dilakukan. Terdapat 2 ujian yang sering digunakan iaitu ujian nisbah kebolehjadian dan ujian Schwartz Bayesian. Bagi kedua-dua ujian ini, parameter model mesti dianggar dengan kaedah kebolehjadian maksimum. Kedua-dua ujian ini juga tidak menguji kebagusan penyesuaian model terhadap data, tetapi menguji sama ada model yang lebih ringkas adalah lebih baik daripada model yang lebih kompleks.

4.1 Ujian nisbah kebolehjadian

Ujian nisbah kebolehjadian menguji hipotesis nol bahawa model A adalah lebih baik daripada model B, dengan model B mempunyai bilangan parameter yang lebih daripada model A. Ujian statistiknya adalah (Klugman et al. 1998),

$$T = 2(l_B - l_A)$$

dengan,

l = logaritma kebolehjadian, lnL, bagi model A dan B yang diuji

Di bawah hipotesis nol, statistik T dianggar bertaburan khi-kuasa dua dengan darjah kebebasan v, di mana v adalah bersamaan dengan perbezaan di antara bilangan parameter model B dengan model A. Nilai T yang besar menunjukkan bahawa model yang lebih kompleks adalah model yang lebih baik.

4.2 Ujian Schwartz Bayesian

Ujian ini memerlukan nilai penalti bagi setiap model yang diuji. Logaritma kebolehjadian, lnL, bagi setiap model perlu ditolak dengan nilai penalti masing-masing agar kelebihan memiliki bilangan parameter yang lebih dan saiz sampel yang lebih besar dapat disingkirkan. Rumus nilai penalti adalah (Schwartz 1978),

$$penalti = r \ln\left(\frac{n}{2\pi}\right)$$

dengan,

r = bilangan parameter model

n = bilangan saiz sampel

Setelah nilai lnL ditolak dengan nilai penalti, nilai ini dipanggil sebagai nilai lnL terubah suai. Nilai lnL terubah suai bagi semua model yang diuji perlu dibandingkan dan model yang mempunyai nilai lnL terubahsuai yang terbesar adalah model yang terbaik.

5. Analisis Data

Sampel data yang diguna adalah sampel tuntutan daripada salah satu produk insurans kesihatan yang ditawarkan oleh salah satu syarikat insurans yang terdapat di Malaysia. Tuntutan ini membayar balik perbelanjaan perubatan sehingga jumlah yang diinsuranskan, termasuklah bilik dan makanan di hospital, rawatan klinik, rawatan perubatan dan rawatan pembedahan. Tuntutan ini boleh dikategorikan kepada jenis insurans perubatan utama. Tempoh tuntutan adalah selama 30 bulan iaitu daripada Januari 1999 sehingga Jun 2001 dan sebanyak 3,734 tuntutan individu telah dilaporkan.

Untuk membina ogif dan histogram bagi sampel data, kesemua 3,734 sampel data telah dibahagikan kepada 6 kumpulan iaitu (0, RM100], (RM100, RM200], (RM200, RM400], (RM400, RM1000], (RM1000, RM5000], (RM5000, •). Nilai-nilai fungsi longgokan empirik, $\tilde{F}_n(x)$, dan fungsi ketumpatan empirik, $\tilde{f}_n(x)$ yang diperoleh untuk setiap kumpulan sampel data diberi di dalam Jadual 1.

Graf ogif dan histogram untuk sampel data berkumpulan telah diperoleh. Graf ogif dan histogram ini ditunjukkan di dalam Rajah 1 dan Rajah 2.

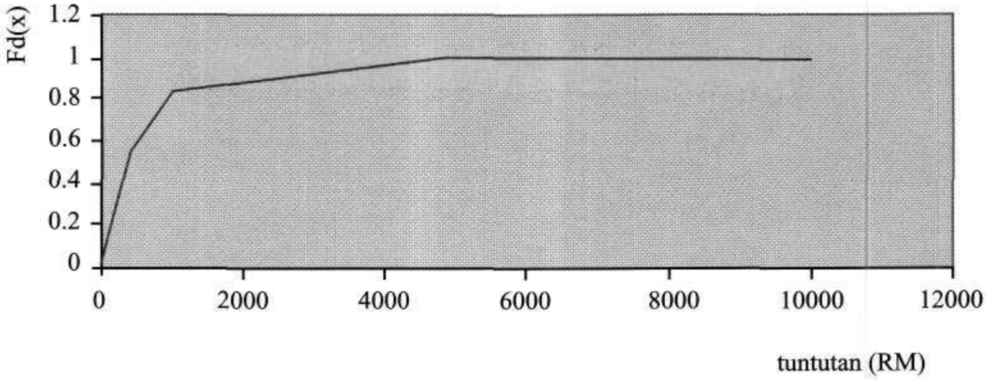
Graf ogif dan histogram boleh memberi beberapa petunjuk kepada jenis taburan yang boleh memodelkan sampel data. Berdasarkan graf ogif dan histogram di atas, taburan yang diperoleh adalah taburan yang terpencong ke kanan (terpencong secara positif) dengan ekor yang sangat berat. Selain itu, fungsi ketumpatan bagi taburan ini menyusut dengan tegas dan mempunyai mod yang bukan sifar.

5.1 Penganggaran berparameter

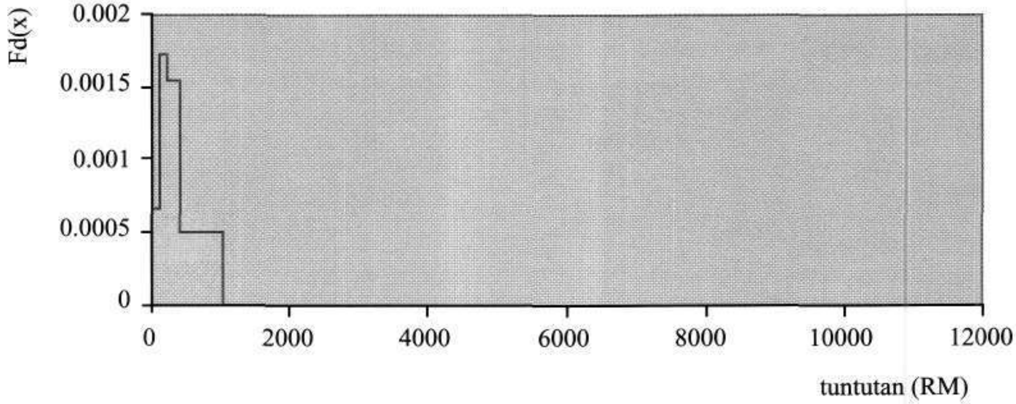
Dengan kaedah kebolehjadian maksimum dan teknik pengoptimuman, anggaran untuk parameter dan nilai logaritma fungsi kebolehjadian yang maksimum bagi setiap taburan telah diperoleh. Nilai-nilai ini diberikan di dalam Jadual 2.

Jadual 1: Nilai-nilai $\widetilde{F}_n(x)$ dan $\widetilde{f}_n(x)$ bagi setiap kumpulan sampel data

kumpulan (RM)	bilangan tuntutan	$\widetilde{F}_n(x)$	$\widetilde{f}_n(x)$
(0,100]	**243**	**0.065078**	**0.000651**
(100,200]	**642**	**0.237011**	**0.001719**
(200,400]	**1149**	**0.544724**	**0.001539**
(400,1000]	**1109**	**0.841725**	**0.000495**
(1000,5000]	**542**	**0.986877**	**0.000004**
(5000,∞]	**49**	**1.000000**	**0.000000**
Jumlah	**3,734**		



Rajah 5.1: Graf ogif untuk sampel tuntutan.



Rajah 5.2: Histogram untuk sampel tuntutan.

Jadual 2: Parameter teranggar dan nilai $\ln L(\underline{\theta})$ yang maksimum bagi setiap taburan

taburan	**bilangan parameter**	**parameter teranggar**	$\ln L(\underline{\theta})$ yang maksimum
1-parameter			
1. eksponen	1	θ=716.79	-28,284.2
2. inverse eksponen	1	θ=252.35	-27,788.6
2-parameter			
3. Pareto	2	α=3.73 θ= 1,878.14	-27,961.2
4. inverse Pareto	2	τ=11.32 θ=24.69	–27,736.4
5. loglogistik	2	γ=1.83 θ=381.66	–27,567.5
6 paralogistik	2	α=1.52 θ=562.68	-27,647
7. inverse paralogistik	2	τ=1.61 θ=253.77	–27,531.2
8. gamma	2	α=0.98 θ=729.15	-28,283.9
9. inverse gamma	2	α=1.36 θ=343.57	–27,555.3
10. Weibull	2	θ=664.69 τ=0.89	–28,222.9
11. inverse Weibull	2	θ=247.57 τ=1.04	-27,782.3
3-parameter			
12. Pareto teritlak	3	θ=153.96 α=1.88 τ=4.14	–27,540.1
13. Burr	3	α=0.61 γ=2.26 θ=265.66	-27,529.9
14. inverse Burr	3	τ=1.79 γ=1.56 θ=230.04	–27,530.1
4-parameter			
15. beta terubahsuai	4	α=0.76 θ=248.20 γ=1.90 τ=1.31	-27,528.8

Berdasarkan keputusan daripada Jadual 2, taburan yang yang mempunyai lnL yang lebih besar adalah taburan yang lebih baik. Oleh itu, taburan yang mempunyai bilangan parameter yang sama dibandingkan dan taburan yang terbaik dipilih bagi setiap kumpulan. Bagi taburan 1-parameter, taburan yang terbaik adalah taburan inverse eksponen (-lnL=-27,788.6, q=252.35), bagi taburan 2-parameter, taburan yang terbaik adalah taburan inverse paralogistik (-lnL=-27,531.2,t=1.61,q=253.77), bagi taburan 3-parameter, taburan yang terbaik adalah taburan Burr (-lnL=27,529.9,a=0.61,g=2.26,q=265.66) dan akhir sekali bagi taburan 4-parameter, taburan yang dipilih adalah taburan beta terubahsuai (-lnL=-27,528.8,a=0.76,q=248.20,g=1.90,t=1.31).

5.2 Ujian penerimaan model

Oleh kerana 4 taburan terbaik bagi kumpulan masing-masing telah dipilih, taburan-taburan ini patut diuji untuk mengetahui kesesuaiannya terhadap sampel data dan untuk memastikan agar mereka boleh diterima sebagai model untuk mewakili sampel data.

5.2.1 Ujian statistik Pearson khi-kuasa dua

Fungsi longgokan untuk kempat-empat taburan yang dipilih diberi di dalam Jadual 3. Untuk ujian ini, sekali lagi kesemua 3,734 sampel data telah dibahagikan kepada 6 kumpulan iaitu (0, RM100], (RM100, RM200], (RM200, RM400], (RM400, RM1000], (RM1000, RM5000], (RM5000, •). Fungsi longgokan untuk keempat-empat taburan ini diguna untuk mendapatkan nilai p_j, E_j, j=1,2,.....,6, dan c^2 masing-masing. Jadual 4 menunjukkan nilai c^2, darjah kebebasan, c pada aras keertian 2.5% dan keputusan sama ada untuk menerima atau menolak H_0 bagi setiap taburan.

Berdasarkan ujian statistik Pearson khi-kuasa dua yang dilakukan ke atas 4 model terbaik, taburan inverse paralogistik (model 2-parameter) dan taburan Burr (model 3-parameter) telah diterima manakala taburan inverse eksponen (model 1-parameter) dan taburan beta terubahsuai (model 4-parameter) telah ditolak.

5.2.2 Ujian statistik Kolmogorov-Smirnov

Bagi ujian ini, data individu akan digunakan. Beza mutlak di antara fungsi longgokan taburan berparameter dengan fungsi longgokan empirik akan dicari pada nilai setiap data individu untuk mendapatkan D_n. Proses ini dilakukan ke atas empat model yang telah dipilih. Fungsi longgokan berparameter untuk keempat-empat model adalah sama seperti yang ditunjukkan di dalam Jadual 3. Jadual 5 menunjukkan nilai D_n, nilai kritikal pada aras keertian 5%, dan keputusan sama ada untuk menerima atau menolak H_0 bagi setiap taburan.

Berdasarkan ujian statistik Kolmogorov-Smirnov pula, sekali lagi taburan inverse paralogistik (model 2-parameter) dan taburan Burr (model 3-parameter) telah diterima manakala taburan inverse eksponen (model 1-parameter) dan taburan beta terubahsuai (model 4-parameter) telah ditolak. Keputusan ini adalah konsisten dengan keputusan yang diperoleh berdasarkan ujian statistik Pearson khi-kuasa dua.

5.3 Ujian pemilihan model

Setakat ini kita mempunyai 2 taburan terbaik, taburan inverse paralogistik (model 2-parameter) dan taburan Burr (model 3-paramter) yang boleh diterima sebagai model untuk mewakili sampel data. Oleh kerana tujuan kita adalah untuk mendapatkan hanya satu model tunggal yang terbaik, ujian untuk pemilihan model bagi model yang mempunyai bilangan parameter yang berbeza perlu dilakukan.

5.3.1 Ujian nisbah kebolehjadian

Ujian ini dilakukan untuk memilih taburan yang lebih baik di antara model yang lebih ringkas, taburan inverse paralogistik, dengan model yang lebih kompleks, taburan Burr. Untuk ujian ini, nilai ujian statistik T adalah,

$$T = 2(-27{,}529.9 - (-27{,}53.2)) = 2.6$$

Nilai kritikal c^2 pada aras keertian 2.5% dan darjah kebebasan 1 adalah 5.02. Oleh itu, H_0 diterima dan model yang lebih ringkas, taburan inverse paralogistik, adalah model yang lebih baik untuk mewakili sampel data.

Jadual 3: Fungsi longgokan untuk taburan inverse eksponen, inverse paralogistik, Burr dan beta terubahsuai

taburan	**fungsi longgokan**
inverse eksponen	$F(x) = e^{-\frac{\theta}{x}}$
inverse paralogistik	$F(x) = \left(\dfrac{\left(\frac{x}{\theta}\right)^{\tau}}{1+\left(\frac{x}{\theta}\right)^{\tau}}\right)^{\tau}$
Burr	$F(x) = 1 - \left(\dfrac{1}{1+\left(\frac{x}{\theta}\right)^{\gamma}}\right)^{\alpha}$
beta terubahsuai	$F(x) = \beta(\tau, \alpha; u), u = \left(\dfrac{\left(\frac{x}{\theta}\right)^{\gamma}}{1+\left(\frac{x}{\theta}\right)^{\gamma}}\right)$

Jadual 4: Nilai c^2, darjah kebebasan, c pada aras keertian 2.5% dan keputusan bagi setiap taburan

taburan	χ^2	Darjah kebebasan	χ^2 pada aras keertian 2.5%	keputusan
inverse eksponen	238.16	4	11.14	Tolak H_0
inverse paralogistik	5.11	3	9.35	Terima H_0
Burr	6.92	2	7.38	Terima H_0
beta terubahsuai	6,067.74	1	5.02	Tolak H_0

Jadual 5: Nilai D_n, nilai kritikal pada aras keertian 5%, dan keputusan bagi setiap taburan

taburan	D_n	Nilai kritikal pada aras keertian 5 %	keputusan
inverse eksponen	0.0869	0.0223	tolak H_0
inverse paralogistik	0.0158	0.0223	terima H_0
Burr	0.0148	0.0223	terima H_0
beta terubahsuai	0.3741	0.0223	tolak H_0

Jadual 6: Nilai penaliti, lnL dan lnL terubahsuai bagi setiap taburan

taburan	lnL	penalti	lnL terubahsuai
inverse paralogistik	-27,531.2	12.775	-27,543.97
Burr	-27,529.9	19.162	-27,549.06

5.3.2 Ujian Schwartz Bayesian

Ujian ini juga dilakukan untuk memilih taburan yang lebih baik di antara model yang lebih ringkas, taburan inverse paralogistik, dengan model yang lebih kompleks, taburan Burr. Nilai penalti, lnL dan lnL terubahsuai bagi kedua-dua model ini diberikan di dalam Jadual 6.

Model yang lebih baik adalah model yang memberikan nilai lnL terubahsuai yang lebih besar. Oleh itu, model yang lebih ringkas, taburan inverse paralogistik, adalah model yang lebih baik untuk mewakili sampel data. Keputusan ini adalah konsisten dengan keputusan yang diperoleh berdasarkan ujian nisbah kebolehjadian.

6. Kesimpulan

Kajian ini bertujuan untuk menyesuaikan suatu taburan berparameter terhadap sampel data tuntutan perbelanjaan perubatan. Model yang terbaik adalah model taburan inverse paralogistik dengan parameter τ=1.61 dan θ=253.77 dan model kedua terbaik pula adalah model taburan Burr dengan parameter α=0.61, γ=2.26 dan θ=265.66.

Pemilihan model dilakukan melalui beberapa peringkat. Dengan kaedah kebolehjadian maksimum, model 1-parameter terbaik, 2-parameter terbaik, 3-parameter terbaik dan 4-parameter terbaik telah dipilih. Kaedah ini memilih model terbaik dengan membandingkan model yang mempunyai bilangan parameter yang sama. Melalui ujian statistik Pearson khi-kuasa dua dan ujian statistik Kolmogorov-Smirnov pula, keempat-empat model telah diuji kesesuaiannya. Hanya 2 model sahaja yang telah diterima. Akhir sekali, dengan ujian nisbah kebolehjadian dan ujian Schwartz Bayesian, hanya 1 model tunggal terbaik sahaja yang dipilih. Ujian ini dilakukan untuk membandingkan model yang mempunyai bilangan parameter yang berbeza.

Setelah model tunggal terbaik diperoleh berdasarkan 'pangkat'nya yang paling tinggi berbanding dengan model yang lain, model ini boleh diaplikasi di dalam insurans. Di antara beberapa aplikasi yang boleh dilakukan adalah, melihat kesan inflasi terhadap model, meramal kebarangkalian terjadinya suatu amaun tuntutan, dengan penggunaan persentil, menjangka tempoh terjadinya suatu amaun tuntutan, dan melihat kesan pengubahsuaian terhadap model dengan mengambil kira faktor had bendungan, deduktibel dan koinsurans.

Rujukan

Brockett, P.L., Cox, S.H., Golany, B., Philips, F.Y., Song, Y. 1995. Actuarial usage of group data: an approach to incorporating secondary data. Society of Actuaries: Transactions. **47**: 89-113.

Chee, H.L. 1990. Health and health care in Malaysia: Present trends and implications for the future. Monograf 3, Institute for advanced studies, Universiti Malaya. Kuala Lumpur.

Hewitt, C.C. & Lefkowitz, B. 1979. Methods for fitting distributions to insurance loss data. Proc. Casualty Actuarial Soc. **66**(125): 139-160.

Hogg, R.V. & Klugman, S.A., 1984. Loss distribution. Canada: John Wiley & Sons.

Jabatan Perangkaan Malaysia. 1996. Buletin Perangkaan Sosial. Kuala Lumpur: Jabatan Perangkaan Malaysia.

Klugman, S.A., Panjer, H.H., Wilmot, G.E., 1998. Loss Models: From data to decision. Canada: John Wiley & Sons.

Philbrick, S.W. 1985. A practical guide to the single parameter Pareto distribution. Proc. Casualty Actuarial Soc. **72**(137): 4-84.

Taylor, G.C. 1994. Modelling mortgage insurance claims experience: a case study. ASTIN Bulletin. **24**(1): 97-129.

Schwartz, G. 1978. Estimating the dimension of a model. Annals of Statistics. **6**: 461-464.

Biography

Noriszura Hj. Ismail is a lecturer in Pusat Pengajian Sains Matematik, Universiti Kebangsaan Malaysia since July 1993. She has graduated from University of Iowa with Bachelor of Science in Actuarial Science (1991) and Master of Science in Actuarial Science (1993). She has wide experience in teaching actuarial science courses including Loss Models, Survival Models, Risk Theory, actuarial Mathematics and Theory of Interests. Her research interests are actuarial modelling, casualty insurance and finance (bond).

Yeoh Sing Yee is working as an actuarial executive in Ing Insurance Company since June 2002. She graduated from Universiti Kebangsaan Malaysia Bachelor of Science in Actuarial Science (2002).